# Supplementary Information for
# "Enhanced Killing of Antibiotic-Resistant Bacteria Enabled By Massively Parallel Combinatorial Genetics"

Allen A. Cheng, Huiming Ding, and Timothy K. Lu

## MATERIALS AND METHODS

**Bacterial Strains and Culture Conditions**. *E. coli* DH5*a* cells used for cloning were obtained from New England Biolabs. *E. coli* ElectroTen-Blue cells used for library electroporation were obtained from Agilent. *E. coli* MG1655 PRO cells (1), a derivative of MG1655, were conjugated with a clinical isolate of *E. coli* producing NDM-1 (CDC Designation 1001728, ATCC) to create MG1655 PRO NDM-1 strains that were used for CombiGEM screening and the time-kill assays (gift from Robert Citorik, Lu lab). *E. coli* EMG2 cells were conjugated with the same NDM-1 clinical isolate to create EMG2 NDM-1, which was used for the phagemid treatment assays (gift from Robert Citorik, Lu lab). Cultures were grown in Luria-Bertoni (LB) broth with the following chemical concentrations as appropriate, unless otherwise noted: aTc at 100ng/mL, carbenicillin (Carb) at 50µg/mL, chloramphenicol (Cm) at 30µg/mL, ceftriaxone at 192µg/mL.

**Chemicals**. T4 DNA ligase, T5 exonuclease, and restriction enzymes were obtained from New England Biolabs. Polymerase chain reactions were performed with HiFi HotStart from Kapa Biosystems. Oligonucleotides were purchased from Integrated DNA Technologies. Antibiotics were purchased from Sigma and Enzo Life Sciences. All other chemicals were purchased from Sigma-Aldrich.

**Minimal Inhibitory Concentration (MIC) Assays.** We performed MIC assays on *E. coli* by broth microdilution with LB broth (Difco Laboratories) using NCCLS methods (2) (Table S1).

**Flow Cytometry**. Each construct representing GFP alone, mCherry alone, or combinations of GFP/mCherry was grown overnight at 37°C and 300rpm in a shaking incubator. Each strain was then diluted 1:100 into three separate cultures for each concentration of aTc and grown for 30 minutes at 37°C and 700rpm in an incubating microplate shaker (VWR). aTc was then added to reach the desired final concentrations in the respective cultures, and the cultures were returned to 37°C and 700rpm for another 3.5 hours. Cultures were then diluted 1:4 in 1xPBS. Fluorescein-isothiocyanate and PE-TexasRed filters on a BD LSRFortessa high-throughput flow cytometer were used to measure GFP and mCherry expression levels, respectively. GFP fluorescence was obtained with a 488nm excitation laser and mCherry fluorescence was obtained with a 561nm excitation laser. 50,000 events were collected for each sample. Gating was performed using FlowJo software (Ashland, OR) and included 20,000 events on average. Geometric means and standard errors of the mean were calculated.

**Library Construction**. To build the barcoded single transcription factor (TF) library, Gibson isothermal assembly (3) was used to combine a unique 6-base pair barcode with the $P_L$tetO promoter, the TF open reading frame (ORF), and the vector backbone. Each barcode was 6 base pairs in length and separated by every other barcode by a genetic distance of at least 2 bases. Transcription factor ORF sequences were obtained from the ASKA Clone(-) library from the National BioResource Project (NIG, Japan)(4). All TF expression constructs were followed by terminator BBa_B0015 from the MIT Registry of Standard Biological Parts. Each barcoded TF contained restriction sites in this configuration: SpeI-Barcode-AvrII-PspOMI-TF expression construct-NotI. Inserts were generated from each vector by PCR with two common primers.

To construct the pairwise combination library, the concentrations of all vectors and inserts were measured on a NanoDrop 2000 (Thermo Scientific) and pooled in equimolar amounts to form a vector pool and an insert pool. The vector pool was digested with AvrII and PspOMI, and the insert pool was digested with SpeI and NotI. Both digests were then purified through phenol/chloroform extraction and ethanol precipitation. The digested vector and insert pools

were ligated together with T4 DNA ligase, purified with phenol/chloroform and ethanol precipitation, and electroporated into ElectroTen-Blue cells per manufacturer's protocol. A total of ~$3*10^6$ transformants were obtained and grown to mid-log in 100mL of LB + 30ug/mL chloramphenicol. Plasmids were isolated through Midi Prep (Qiagen) and electroporated into *E. coli* MG1655 PRO NDM-1.

**Antibiotic Treatments for CombiGEM Analysis**. Antibiotic assays of CombiGEM libraries for analysis with high-throughput sequencing were performed with *E. coli* MG1655 PRO NDM-1 containing the pairwise CombiGEM library at 37°C in 96-well clear-bottom plates on a VersaMax Microplate Reader/Shaker (Molecular Devices). All culture wells held 200µL. Frozen stocks of *E. coli* MG1655 PRO NDM-1 with the CombiGEM libraries were diluted into LB with or without 100ng/mL aTc. When these cultures reached an OD600 = 0.6, they were diluted 1:100 into wells with or without antibiotic and with or without 100ng/mL aTc. Antibiotic concentrations used were: ceftriaxone low: 64µg/mL; ceftriaxone high: 256µg/mL; imipenem low: 32µg/mL; imipenem high: 96µg/mL; piperacillin-tazobactam low: 64µg/mL piperacillin, 8µg/mL tazobactam; piperacillin-tazobactam high: 256µg/mL piperacillin, 32µg/mL tazobactam; gentamicin low: 32µg/mL; gentamicin high: 256µg/mL. Each unique culture condition was grown in duplicate. These wells were grown until OD600 = 0.3 (early logarithmic phase) or 0.9 (late logarithmic phase) with shaking. All DNA from well cultures was harvested by alkaline lysis and ethanol precipitation. Each miniprepped sample was assayed for concentration via SYBR Fast qPCR kits (Kapa Biosystems).

**High-Throughput Sequencing**. Each sample was prepared for Illumina HiSeq sequencing by adding an indexing barcode and Illumina anchor sequences via PCR using primers indicated below. To prevent PCR bias that would skew the population distribution, PCR reactions were terminated during exponential phase. PCR products were purified with AMPure XP beads (Agencourt), concentration quantified via qPCR, and pooled in equimolar amounts. Multiplexed samples were then sequenced using the primers indicated below.

**Primers for Amplifying CombiGEM Populations for Sequencing.**

Forward primer: AATGATACGGCGACCACCGAGATCTACACCGCTGGCAAGTGTAGC

Barcoded reverse primer:
CAAGCAGAAGACGGCATACGAGATNNNNNNGGGAGGGCCCGTTG

where NNNNNN denotes a specific barcode unique for each experimental sample.

Illumina sequencing primer:
CCACGAGGATTCGAAAAGGTGAACCGACCCGGTCGATGCACTAGT

Illumina indexing primer: CCTAGGAGCAAGTACGAACAACGGGCCCTCCC

**S-Score Population Analysis.** Raw reads for each gene pair in each experiment were processed from sequencing data. To ensure valid log transformation downstream, a pseudo read of 1 was added for each gene pair-experiment combination. Reads for each combination were normalized to the total reads in each experiment and to fluorophore controls (pairwise combinations consisting only of eCFP, GFP, and mCherry) in each experiment. To correct for the batch effect of aTc using the multiplicative model, we divided the normalized reads for each gene pair-experiment by the median reads of that gene pair from all experiments in either aTc on or off. Finally, the normalized and batch corrected reads were log transformed, producing the relative abundance for each combination of drug and gene pair. The difference between the relative abundances of a gene pair in two experiments corresponds to the log ratio of their normalized and batch corrected reads. The final abundance for each combination of drug and gene pair is the average of the Z-score-transformed relative abundances of two duplicates. The calculation of the S-score is based on the additive model between the relative abundance for a gene pair under a particular drug condition and the mean of abundances for a reference set that consists of all the drug screens. The formula for the interaction S-score is defined as $S_{ij}^{k} = \frac{\alpha_{ij}^{k} - \overline{\alpha}_{ij}}{\sigma_{ij}}$, where *i,j* is a

combination of genes *i* and *j*, *k* spans a set of experiments, $\bar{\alpha}_{ij} = \frac{1}{n_{ij}} \sum_k \alpha_{ij}^k$ is the mean of abundances $\alpha_{ij}^k$ from $n_{ij}$ experiments, and $\sigma_{ij}^k$ is the overall standard deviation with a minimum bound that accounts for the systematic variance as well as the variance of abundances from duplicates. In all of the heatmaps representing S-score data (Fig. 1d, Fig. S5, and Fig. S7), any S-scores that were greater than 2.4 or less than -2.4 are represented by the strongest green and red shading, respectively. For the data in Fig. 1d and Table 1, we randomly selected hits from the 200 top-scoring combinations and controls with neutral S-scores based on the high ceftriaxone (256µg/mL) condition, combining both the early log and late log data.

**Log-Ratio Analysis.** Raw reads for each gene pair in each experiment were processed from sequencing data. To ensure valid log transformation downstream, a pseudo read of 1 was added for each gene pair-experiment combination. Reads for each combination were normalized to the total reads in each experiment. A log-ratio of normalized reads comparing aTc against no aTc was calculated for each beta-lactam antibiotic at high concentrations in each growth stage, for a total of eight ratios (no antibiotic, ceftriaxone, imipenem, piperacillin-tazobactam at early log and late log). Combinations were sorted by the mean log ratio across all antibiotic-time conditions. For the data in Table 1, we randomly selected hits from the 200 top-scoring combinations based on the high ceftriaxone (256µg/mL) condition, combining both the early log and late log data.

**Clustering Analysis.** Hierarchical clustering was performed across a subset of gene pairs and experiments with Euclidean correlation and complete linkage.

**Antibiotic Time-Kill Assays.** Combinations identified through the analysis methods described above were generated into the vector backbone, verified through Sanger sequencing, and transformed into *E. coli* MG1655 PRO NDM-1. Frozen stocks were grown overnight in LB + Carb + Cm, then diluted 1:100 into 2mL LB + Cm, and grown at 37°C and 300rpm for 1 hour in

a shaking incubator. aTc and ceftriaxone were added where appropriate to reach final concentrations of 100ng/mL and 192µg/mL, respectively, and cultures were returned to 37°C and 300rpm. To obtain colony-forming-unit (CFU) counts at indicated time points, 100µL of relevant cultures were collected, washed in 1x phosphate-buffered saline (PBS), and resuspended in 100µL of PBS. Each resuspension was serially diluted with PBS, and 10µL of each dilution was plated in duplicate on drug-free LB agar plates. LB agar plates were then incubated at 37°C overnight, and colony numbers were counted. Three biological replicates were performed for each combination and culture condition. Error bars represent standard error of the mean.

**Phagemid Generation.** Phage particles containing combinatorial constructs on phagemids were generated using a helper phage-free system as previously described (5). M13cp was modified to confer kanamycin resistance through the MluI restriction site and transformed into DH5α. This strain was then made chemically competent, transformed with the combinatorial constructs described, and plated on LB-kanamycin-chloramphenicol plates. Colonies from each transformation were picked and grown overnight in LB-kanamycin-chloramphenicol at 37°C and 300rpm. Phage particles were collected from supernatants by centrifugation at 3200g for 30min and filter sterilized. Levels of phage production were determined by the titration of lysates on *E. coli* EMG2 NDM-1 cells plated on LB and LB-chloramphenicol plates.

**Phagemid Treatment**. An *E. coli* EMG2 NDM-1 culture was grown overnight at 37°C in LB-carbenicillin. The following day, the culture was diluted 1:500 into 1mL of LB and re-grown at 37°C for one hour and 300rpm. Phage particles containing phagemids for combinatorial gene expression were added to each culture tube at an MOI (multiplicity-of-infection) of 10:1 and cultured at 37°C for 30 minutes and 300rpm. aTc and ceftriaxone were then added where appropriate, which was designated as time t = 0 in Fig. 4. 100µL aliquots from cultures were collected at indicated time points, washed in 1x PBS, and resuspended in 100µL of PBS. Serial dilutions were performed with PBS, and 10µL of each dilution was plated in duplicate on drug-free LB agar plates. LB agar plates were incubated at 37°C overnight before counting. Three

biological replicates were performed for each combination and culture condition. Error bars represent standard error of the mean.

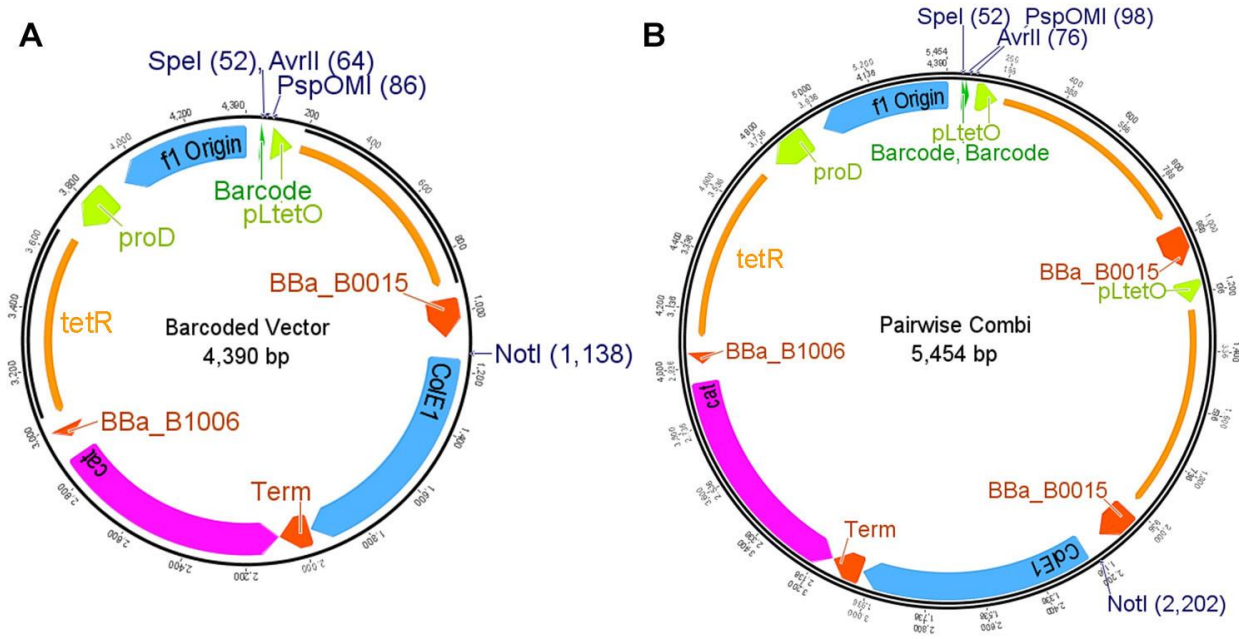


**Figure S1.** Combinatorial expression plasmid design, which contains a ColE1 origin, an f1 origin, and a chloramphenicol resistance gene (*cat*). **(a)** A single gene expression construct (GFP here) is barcoded and flanked by restriction sites for CombiGEM library assembly. **(b)** A pairwise combination is constructed through the ligation of a vector digested with AvrII/PspOMI and an insert digested with SpeI/NotI. Barcodes are in close proximity for sequencing. Restriction sites are retained for further iterations to construct higher-order combinations.

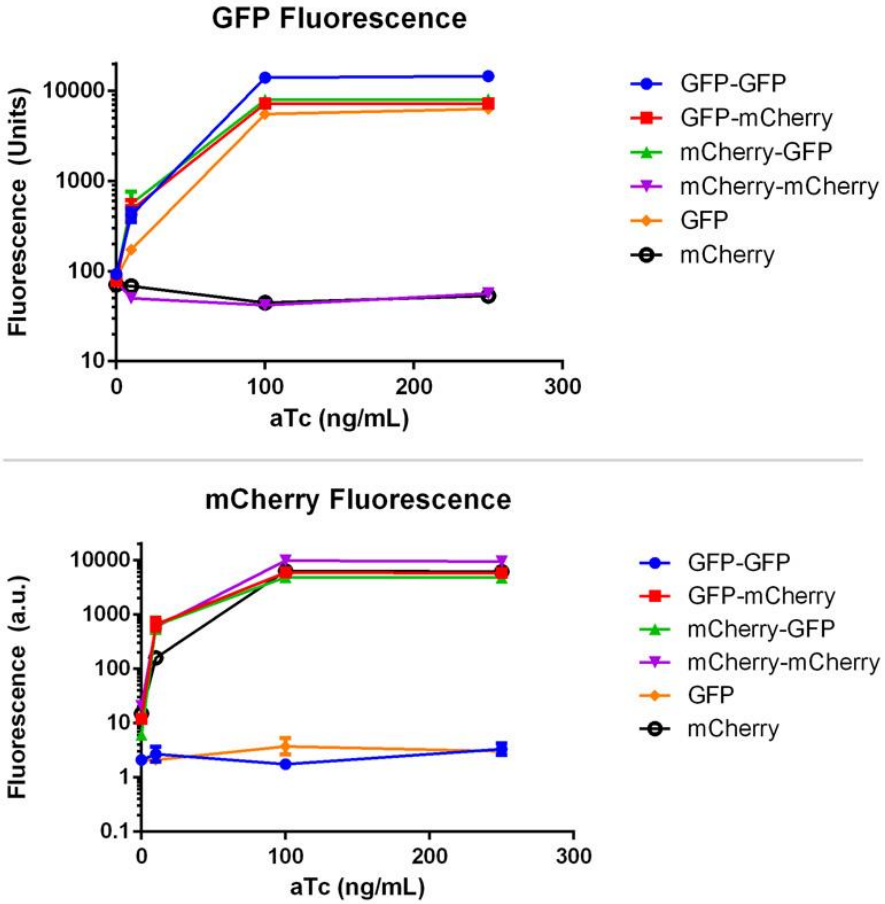


**Figure S2.** Fluorescence induction curves. Pairwise combinations of GFP and mCherry and single GFP and mCherry constructs were induced with aTc, and fluorescence was measured by flow cytometry 3.5 hours after induction. The 250ng/mL data points shown here are reproduced in Figure S3a for clarity.

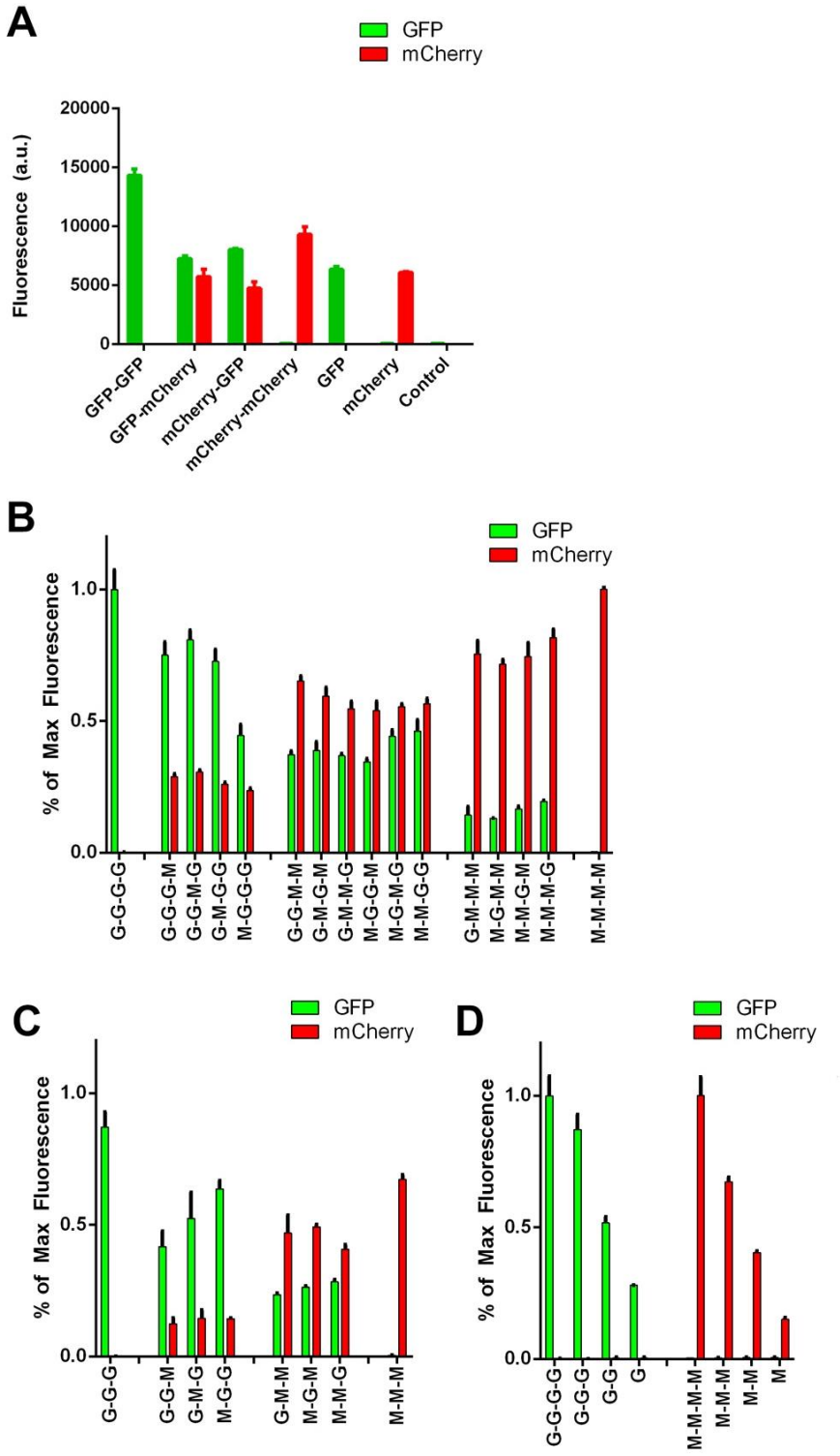


**Figure S3.** Gene expression from high-order CombiGEM constructs. All constructs were induced with aTc at 250ng/mL and assessed by flow cytometry after 3.5 hours of growth. Labels:

G = GFP, M = mCherry. Values from the same construct are repeated across this figure. **(a)** All four pairwise combinations of GFP and mCherry and single GFP and mCherry constructs are compared. GFP-mCherry and mCherry-GFP showed comparable levels of fluorescence, indicating the minimal influence of gene order on expression level. GFP-GFP and mCherry-mCherry showed approximately double expression levels of GFP and mCherry, respectively, compared to heterogeneous combinations and single constructs. The negative fluorescence control shown is an expression construct with *caiF* + *modE*. **(b-d)** Fluorescence levels for GFP and mCherry are shown as the % of max fluorescence for each fluorophore across all experiments shown. **(b)** Four-wise combinations and **(c)** three-wise combinations of GFP and mCherry show comparable fluorescence levels regardless of order. **(d)** Fluorescence is additive with increasingly higher-order combinations.

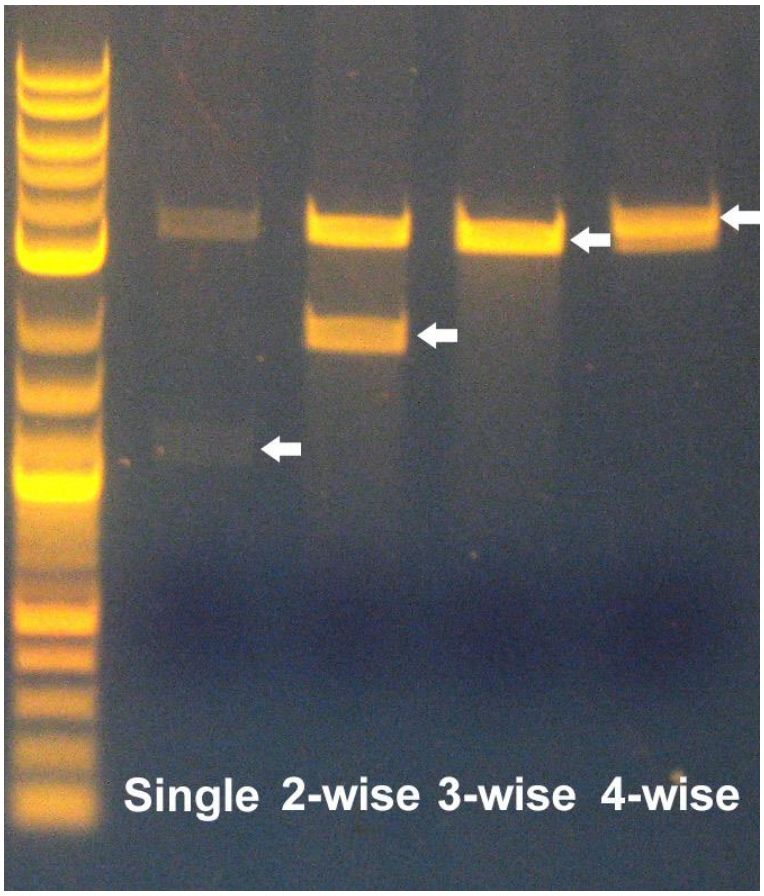


**Figure S4.** Constructing higher-order combinations through CombiGEM. The barcoded GFP vector was iteratively digested and ligated with barcoded GFP inserts to yield 2-wise, 3-wise, and 4-wise combinations of GFP expression cassettes. Vectors were then digested with SpeI and NotI to separate the insert region from the static vector for visualization of successful DNA assembly, as shown here. Agarose gel electrophoresis shows the progressive size increase of the insert with each iteration (arrows), compared to 2-log ladder (New England Biolabs).

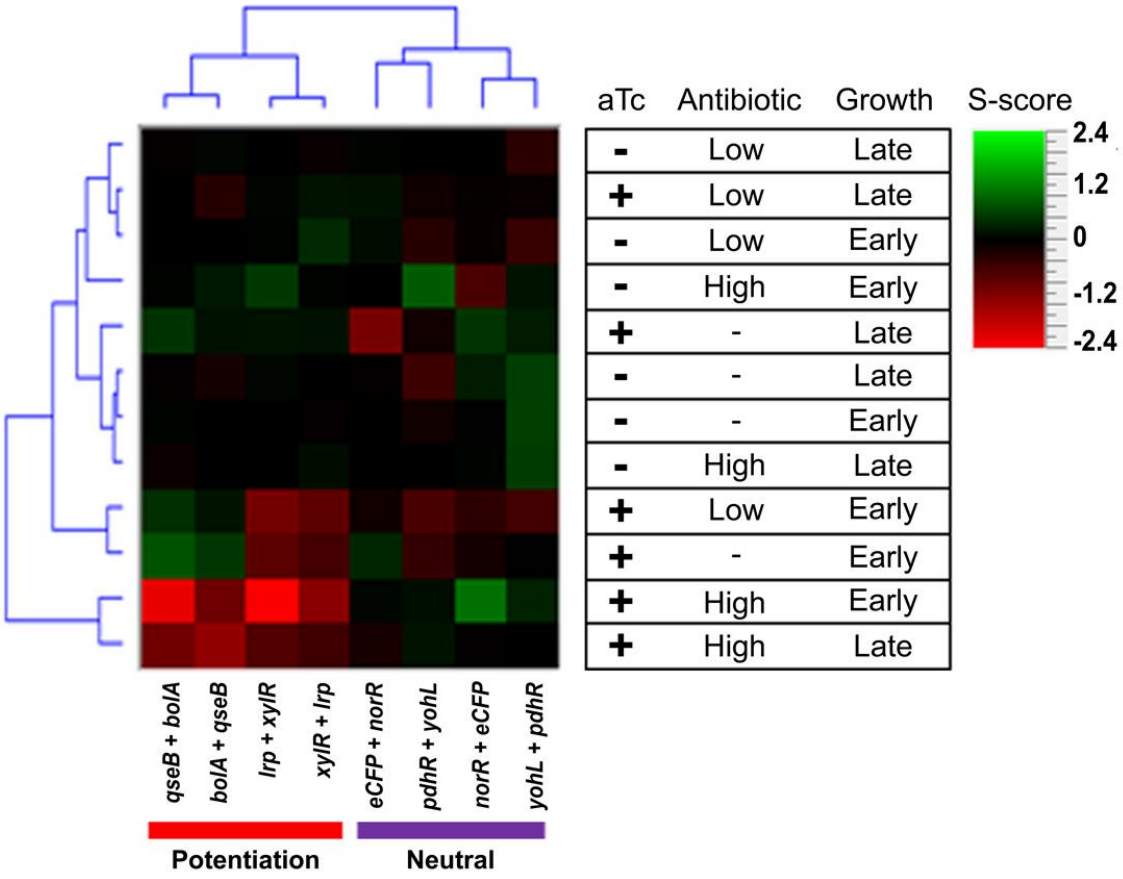


**Figure S5.** Hierarchical clustering of specific gene combinations and their reciprocals, which have the same gene combinations but in reverse order within the expression construct. The antibiotic used in this experiment was ceftriaxone, which was applied in low and high concentrations as described in the Methods. Gene pairs conferring the same phenotype and their reciprocals clustered together. This data is a superset of the data shown in Fig. 1d.

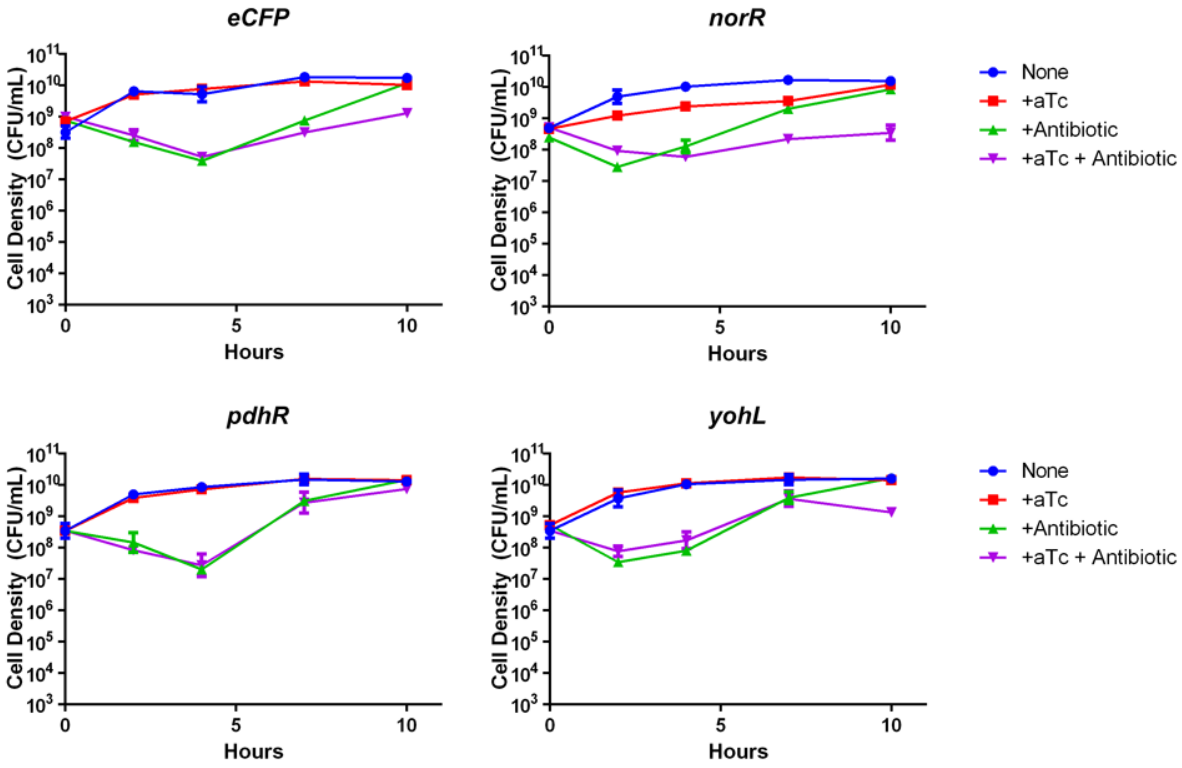


**Figure S6.** Time-kill curves of NDM-1 with individual constituent genes from the control combinations shown in Fig. 2b. Individual genes show minimal potentiation of ceftriaxone. Error bars represent SEM (small error bars are obscured by symbols).

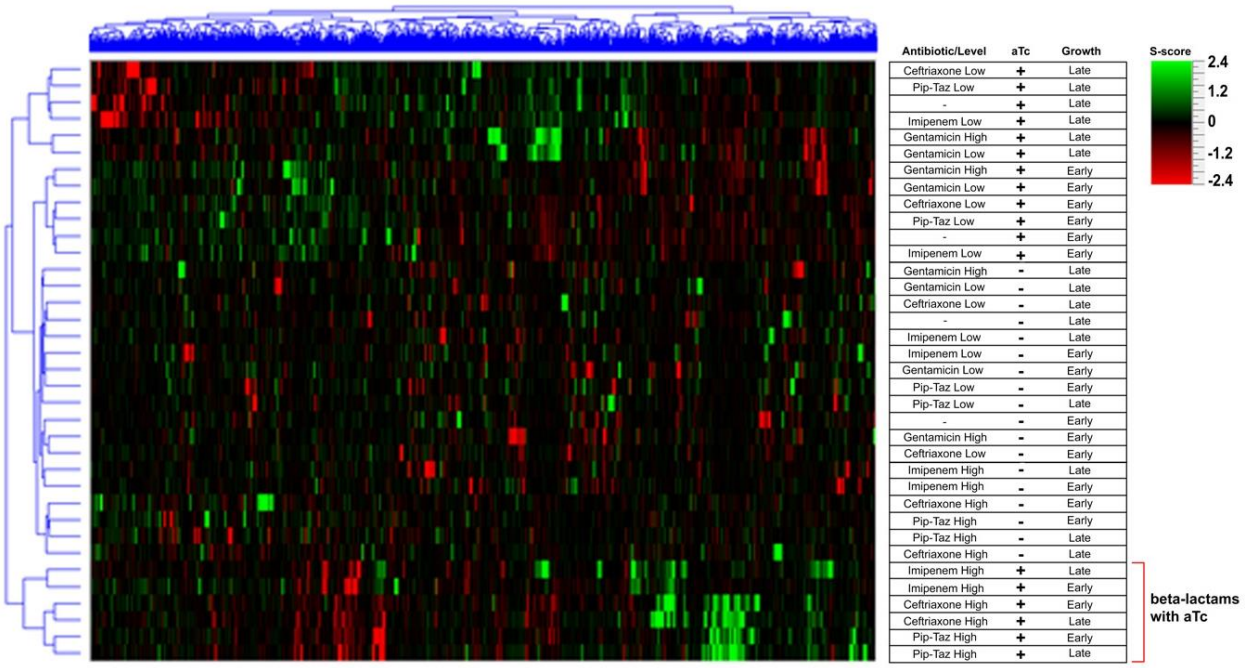


**Figure S7.** Hierarchical clustering of significant hits across all antibiotic-aTc conditions. Antibiotics were applied in low and high concentrations, as described in the Methods. Cells were collected for analysis in early logarithmic and late logarithmic growth phases, as described in the Methods. 6311 gene pairs show at least one S-score above 2.4 or below -2.4 in any experiment, where 2.4 is three times the standard deviation of the S-score population distribution. Euclidean correlation with complete linkage was employed in clustering analysis. Hits for the beta-lactams imipenem, ceftriaxone, and piperacillin-tazobactam (Pip-Taz), in the presence of aTc, cluster together (red bracket), separately from gentamicin.

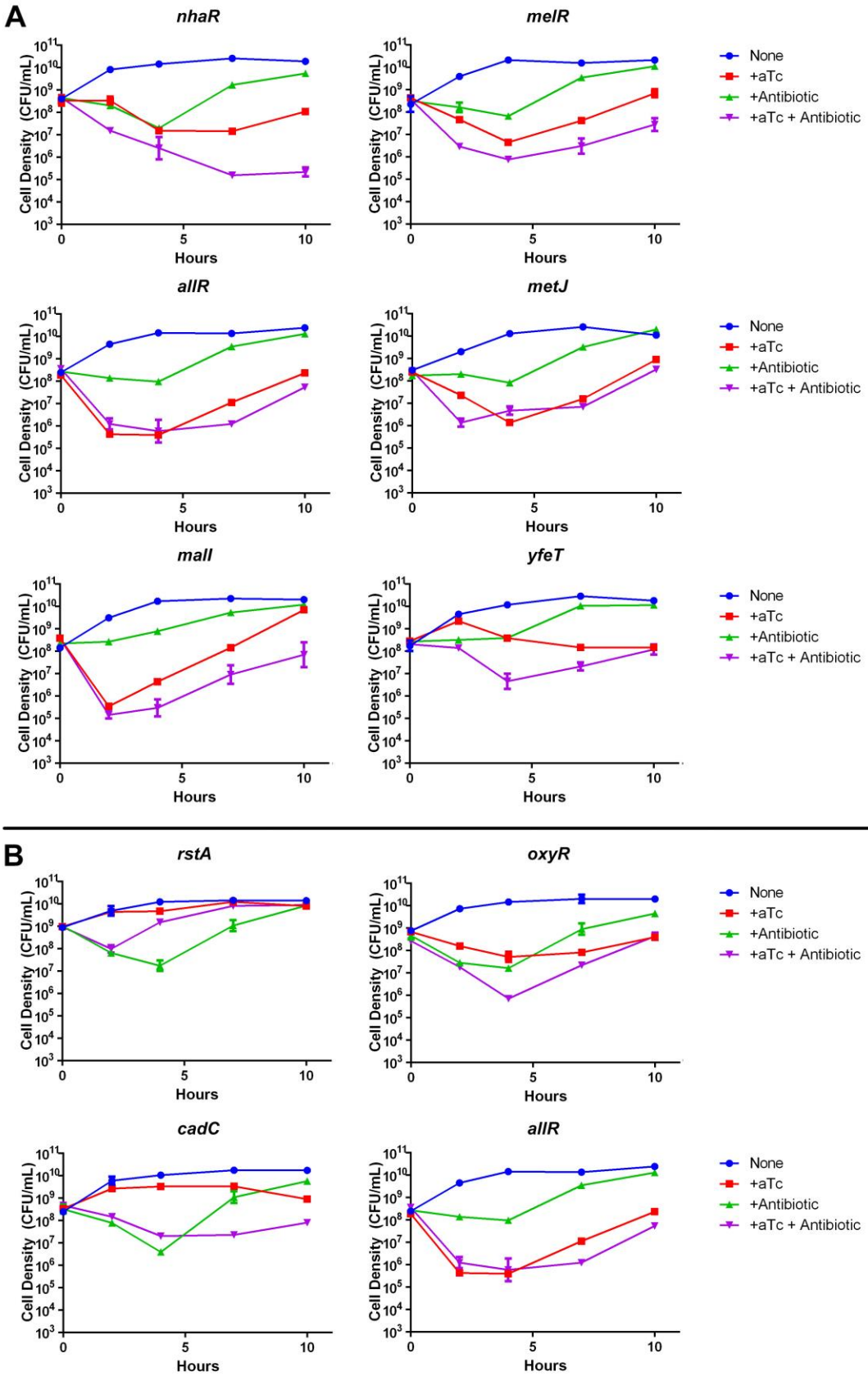
A
nhaR
melR
allR
metJ
mall
yfeT
B
rstA
oxyR
cadC
allR
Cell Density (CFU/mL)
Hours
None
+aTc
+Antibiotic
+aTc + Antibiotic

**Figure S8.** Time-kill curves of NDM-1 with individual constituent genes from the combinations shown in Fig. 3. **(a)** Individual genes from lethal combinations in Fig. 3a show weaker potentiation of ceftriaxone activating than their corresponding pairwise combinations. Furthermore, neither gene within each combination shows a dominating effect compared with each other or the combination of the two genes. **(b)** Individual genes from the control combinations in Fig. 3b. The time-kill curve for *allR* is duplicated from a) for ease of reference. Error bars represent SEM (small error bars are obscured by symbols).

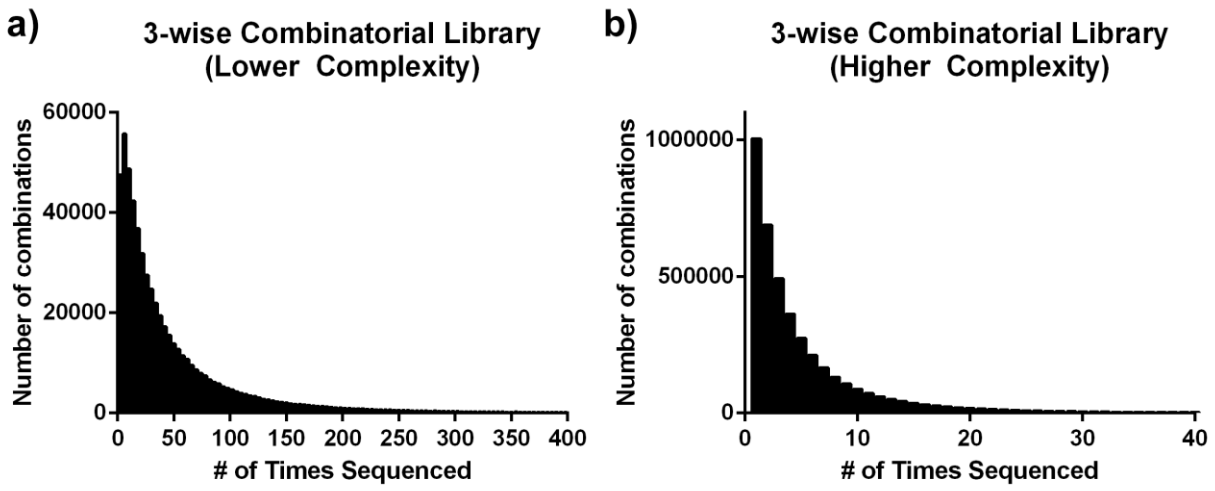


**Figure S9.** Construction and censusing of 3-wise combinatorial TF expression libraries with Illumina sequencing. **(a)** A lower complexity 3-wise library consisted of 189 x 187 x 18 genes (636,174 possible combinations). From 30 million sequencing reads, we recovered 582,433 unique combinations (92%). **(b)** A higher complexity 3-wise library consisted of 189 x 187 x 187 genes (6,609,141 possible combinations). From 20 million reads, we recovered 4,042,316 combinations (61%).

**Table S1.** NDM-1 Minimal Inhibitory Concentrations (MICs) (μg/mL)

| **Antibiotic** | **Tested in *E. coli* MG1655 PRO NDM-1** | **Previously Reported in *Klebsiella pneumoniae* NDM-1 (6)** | **Previously Reported in *E. coli* NDM-1 (7)** | **Tested in *E. coli* MG1655** |
|---|---|---|---|---|
| Amoxicillin | >256 | N/A | >512 | <2 |
| Cefepime | N/A | >32 | 64 | N/A |
| Ceftriaxone | 256 | >32 | N/A | <2 |
| Colistin | <0.125 | 32 | Susceptible | <0.125 |
| Gentamicin | >256 | >10 | N/A | <2 |
| Imipenem | >8 | 8 | 6 | <0.5 |
| Piperacillin-Tazobactam* | >256 | >128 | 512 | <1 |

*Piperacillin-tazobactam was administered in an 8:1 ratio. MIC value indicates the concentration of piperacillin.

**Table S2.** Plasmids used in this work.

| **Plasmid ID** | **Description** | **Use** |
|---|---|---|
| pAAC001 | ColE1 origin, Cm resistance, *tetR* expression | Base Plasmid to Build Library Members |
| pAAC002 Library | pAAC001 with pL(tetO) promoters expressing TF combinations | Pairwise Combinatorial Library |
| pAAC003 | pAAC001 with pL(tetO) promoters expressing *qseB* + *bolA* | Potentiating Gene Pair |
| pAAC004 | pAAC001 with pL(tetO) promoters expressing *lrp* + *xylR* | Potentiating Gene Pair |
| pAAC005 | pAAC001 with pL(tetO) promoters expressing *eCFP* + *norR* | Neutral Potentiating Gene Pair |
| pAAC006 | pAAC001 with pL(tetO) promoters expressing *pdhR* + *yohL* | Neutral Potentiating Gene Pair |
| pAAC007 | pAAC001 with pL(tetO) promoters expressing *torR* + *metR* | Lethal Gene Pair |
| pAAC008 | pAAC001 with pL(tetO) promoters expressing *nhaR* + *melR* | Lethal Gene Pair |
| pAAC009 | pAAC001 with pL(tetO) promoters expressing *allR* + *metJ* | Lethal Gene Pair |
| pAAC010 | pAAC001 with pL(tetO) promoters expressing *malI* + *yfeT* | Lethal Gene Pair |
| pAAC011 | pAAC001 with pL(tetO) promoters expressing *rstA* + *oxyR* | Control Lethal Gene Pair |
| pAAC012 | pAAC001 with pL(tetO) promoters expressing *cadC* + *allR* | Control Lethal Gene Pair |

## REFERENCES


1. Lutz R & Bujard H (1997) Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res* 25(6):1203-1210.
2. National Committee for Clinical Laboratory Standards (2000) Methods for dilution antimicrobial susceptibility tests for bacteria that grow aerobically, fifth ed. Approved standard M7-A5.
3. Gibson DG, *et al.* (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature methods* 6(5):343-345.
4. Kitagawa M, *et al.* (2005) Complete set of ORF clones of Escherichia coli ASKA library (a complete set of E. coli K-12 ORF archive): unique resources for biological research. *DNA Res* 12(5):291-299.
5. Chasteen L, Ayriss J, Pavlik P, & Bradbury AR (2006) Eliminating helper phage from phage display. *Nucleic acids research* 34(21):e145.
6. Mochon AB, *et al.* (2011) New Delhi metallo-beta-lactamase (NDM-1)-producing Klebsiella pneumoniae: case report and laboratory detection strategies. *Journal of clinical microbiology* 49(4):1667-1670.
7. Poirel L, Lagrutta E, Taylor P, Pham J, & Nordmann P (2010) Emergence of metallo-beta-lactamase NDM-1-producing multidrug-resistant Escherichia coli in Australia. *Antimicrobial agents and chemotherapy* 54(11):4914-4916.